

Wine quality classification

Final Presentation

Miroslav Mažgut (525136), Filip Gregora (525265)

Faculty of Informatics, Masaryk University

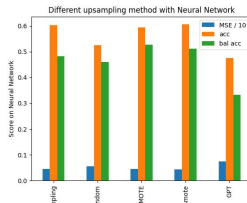
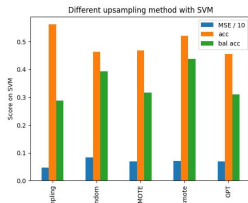
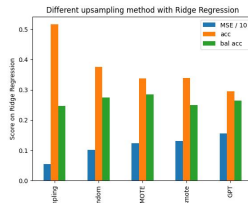
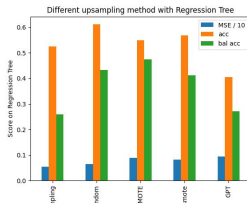
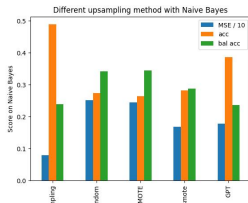
May 15, 2025

What was implemented

- ML Models
 - Naive Bayes
 - Regression Tree
 - Ridge Regression
 - SVM
 - Neural Network
- Upsampling:
 - Random sampler
 - SMOTE
 - KMeanSMOTE
 - Large language model (prompting ChatGPT)
- Training model on upsampled data and evaluate them
- Trimming border classes (class 3 and class 9)

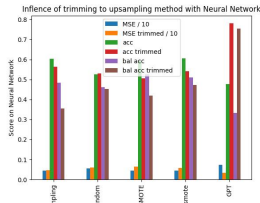
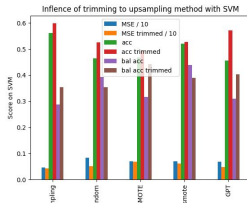
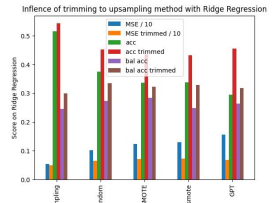
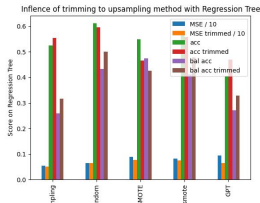
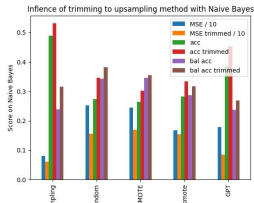
Upsampling Results

- No upsampling has the lowest RMSE
- Upsampling increase accuracy and balanced accuracy
- GPT upsampling shows poor results



Trimmed Upsampling Result

■ Trimmed categories have better results



Upsampling results for Neural Network

Upsampling	RMSE	accuracy	bal. accuracy
no upsampling	0.4561	0.6038	0.4832
random	0.561	0.5254	0.461
SMOTE	0.4495	0.5938	0.5267
KMsmote	0.4421	0.6062	0.5114
GPT	0.7393	0.4762	0.3335

Table: Results on upsampled dataset

Upsampling	RMSE	accuracy	bal. accuracy
no upsampling	0.4699	0.5638	0.3549
random	0.5999	0.5313	0.4525
SMOTE	0.638	0.505	0.419
KMsmote	0.5735	0.5422	0.4718
GPT	0.3384	0.7804	0.7545

Table: Results on upsampled trimmed dataset

Division work

Filip Gregora: Model training, Model evaluation

Miroslav Mažgut: Model evaluation, Trimmed datasets

Thank You for Your Attention!

MUNI

FACULTY

OF INFORMATICS